3sum

5 December, 2020 10:43

3sum:

Given sets $A, B, C \subseteq \{-U, \dots U\}$ decide

$$\exists a \in A, \exists b \in B, \exists c \in C \quad \text{s.t.} \quad a+b+c=0?$$

Alg.:
- $O(n^3)$ .. trivial
- $O(n^2 \lg n)$ ... easy - sort C & look-up $-(a+b)$ for each $a \in A$, $b \in B$
- $O(n^2)$ ... using hashing
- $O(n^2 \cdot \frac{\lg\lg n}{\lg^2 n})$ ... using bit tricks

Variants:

1) sets A, B, C $\quad \exists a,b,c \quad a+b+c=0?$
2) sets A, B, C $\quad \exists a,b,c \quad a+b=c?$

$C = -C$

3) sets A, B, C, int t $\quad \exists a,b,c \quad a+b+c=t?$ $\quad C \to C-t$
4) set X $\quad \exists x+y+z \in X \quad x+y+z=0?$

"$\Rightarrow$" $\quad A, B, C := X$

"$\Leftarrow$" $\quad X = \underbrace{A+4U}_{\{a+4U;\, a \in A\}} \cup B+8U \cup C-12U$

solution $x, y, z \in X$, $x+y+z=0$ $\Rightarrow$ solution $a \in A, b \in B, c \in C$, $a+b+c=0$

no solution $x+y+z=0$ can contain two elements originating in one of these sets $A+4U$, $B+8U$, $C-12U$

Alg.: deterministic, $O(n^2)$, variant with a single set A, $a+b+c=0$.

$\leftarrow i$

$\boxed{\quad \leq \quad A \quad \leq \quad}$

$j \rightarrow$

1) sort A in increasing order

1) Sort A in increasing order
2) for each $c \in A$
   set $i := n$; $j := 1$
   while $i > 0$ & $j \leq n$
      if $a_i + a_j = -c$ return (a_i, a_j, c)
      if $a_i + a_j > -c$ $i := i-1$
      if $a_i + a_j < -c$ $j := j+1$
3) return no solution.

For each a_i, there exists a unique j_i
s.t. $a_i + a_{j_i} < -c$ &
$a_i + a_{j_i+1} \geq -c$ &
$a_i + a_{j_i+2} > -c$

as i decreases, j increases.

Alg: For small U, $O(n + U \lg U)$, $A, B, C \subseteq \{-U, \dots, U\}$
$a+b+c=0$.

1) define polynomials

$$p_A(x) = \sum_{a \in A} x^{U+a} \qquad p_B(x) = \sum_{b \in B} x^{U+b} \qquad p_C(x) = \sum_{c \in C} x^{U+c}$$

2) Compute $p_A(x) \cdot p_B(x) \cdot p_C(x) = p_{ABC}(x)$

The coeficient x^{3U} in $p_{ABC}(x)$ is equal to the # of triples $(a,b,c) \in A \times B \times C$ s.t.
$a+b+c=0$.

3) return the coeficient of x^{3U}.

the multiplication of $p_A(x)$ by $p_B(x)$ & $p_C(x)$ can be done in time $O(U \lg U)$ using FFT.

(Multiplication of two poly's of degree $\leq d$ represented by their coef's can be done in time $O(d \lg d)$)

k-sum:

Given sets $X_1, X_2, \dots X_k \in \{-u, \dots u\}$, decide
$\exists x_1 \in X_1, \exists x_2 \in X_2, \dots \exists x_k \in X_k$ $\quad x_1 + x_2 + \dots + x_k = 0$?

1) $O(n^{k-1} \lg n)$ - alg. ... "trivial"

2) $\hat{O}(n^{\lceil k/2 \rceil})$ - algorithm: take $\lceil k/2 \rceil$-tuples from $X_1 \dots X_{\lceil k/2 \rceil}$ $\to A = \{x_1 + x_2 + \dots + x_{\lceil k/2 \rceil};\ x_i \in X_i\}$

& $\lfloor k/2 \rfloor$-tuples from $X_{k-\lceil k/2 \rceil} \dots X_k \to B = \{x_{k-\lceil k/2 \rceil} + \dots + x_k,\ x_i \in X_i\}$

solve $\exists a \in A\ \exists b \in B\quad a + b = 0$?

↳ $O((|A| + |B|) \cdot \lg n)$ time

$\to O(n^{\lceil k/2 \rceil} \cdot \lg n)$.

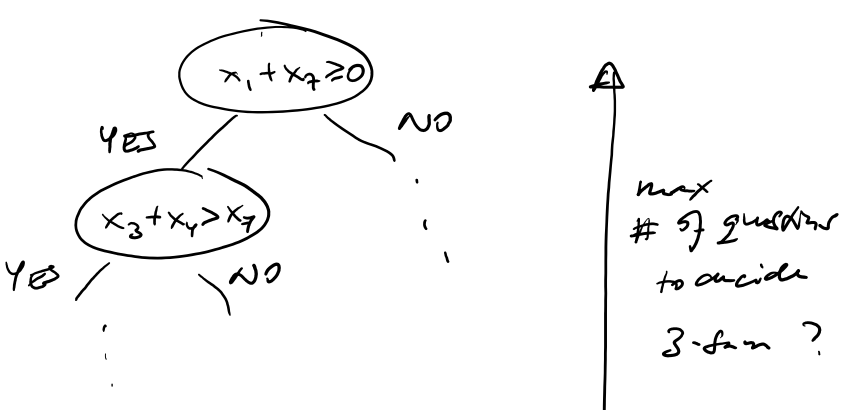
Decision tree complexity

- want to estimate, how many comparisons of elts one has to make, to solve 3-sum.

we allow arbitrary comparisons of the form

$\sum c_i x_i > 0$ where c_i's are fixed coefficients & x_i are elts in X indexed by their index in X.

decision tree



3-Sum ?

output (YES) (NO)

<u>Ex</u>: sorting in comparison model (= $\exists$ dec. tree) requires $\Omega(n \lg n)$ questions (= comparisons).

<u>Thm</u>: 3-Sum has decision tree of depth $O(n^{3/2} \lg n)$

$\to$ alg. in time $O\left(n^2 \frac{\text{poly}\lg\lg n}{\sqrt{\lg n}}\right)$ no bit-tricks

$g = \sqrt{n}$

<u>Alg</u>:

1) sort A in increasing order
2) partition A into n/g groups $A_1, \dots A_{n/g}$ of size at most g

   $\max A_i \leq \min A_{i+1} \quad \forall i < \frac{n}{g}$

3) sort $D = \bigcup_{i=1}^{n/g} A_i - A_i = \{a-b; \exists i, a, b \in A_i\}$

4) For all i, i', sort $A_{i,i'} = A_i + A_{i'} = \{a+b;\ a \in A_i, b \in A_{i'}\}$

5) for each $c \in A$ // check $\exists a, b \in A$ s.t. $a+b=c$

   set $i = \frac{n}{g}$, $j = 1$

   while $i > 0$ & $j \leq \frac{n}{g}$

      if $-c \in A_{i,j}$ return "solution exists"

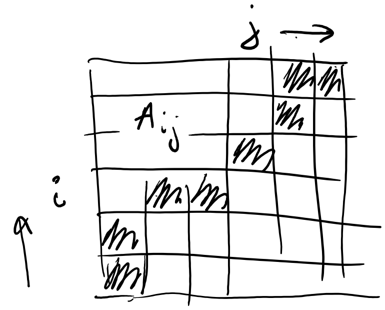
      if $\min(A_i) + \max(A_j) > -c$; $i = i-1$

      else $j = j+1$

6) return "no solution"

$i \to$

6) return "no solution"



- boxes where $-c$ could possibly be
- algorithm enumerates over them

\# of comparisons

1) sort A   3) sort D   4) sort $A_{i,i'}$ 5)

$$O(n \lg n) + O(ng \lg(ng)) + \cancel{O} + O\Big(\underbrace{n}_{\#\text{ of } c \in A} \cdot \underbrace{\frac{n}{g}}_{\text{while loop iterations}} \cdot \underbrace{\lg(g^2)}_{\text{bin search in } A_{i,j}}\Big)$$

to decide whether

$$\underset{\in A_{i'}}{\overset{\in A_i}{a+b}} < \underset{\in A_{i'}}{\overset{\in A_i}{c+d}}$$ we check in sorted D $$\overset{\in D}{a-c} < \overset{\in D}{b-d}$$

think of D as containing triples of indices (i,j,k)

$i \in \frac{n}{g} \quad j,k \in g$

j^{th} & k^{th} elt from A_i

$\to a_j - a_k \in D$

<u>total</u>: $O\left(\left(ng + \frac{n^2}{g}\right)\lg ng\right) = O\left(n^{3/2}\lg n\right)$ for $g = \sqrt{n}$. ☒

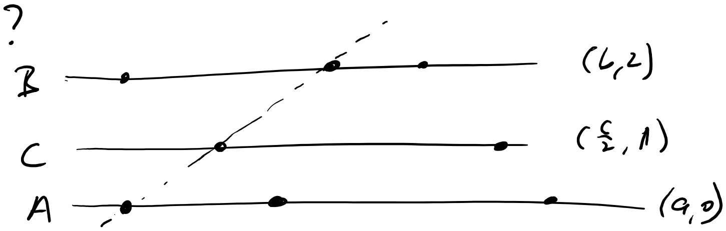
3-sum-hard problems

<u>Geom Base</u>

... horizontal lines

Geom Base

Given $3 \times n$ points on three horizontal lines, are any three of them colinear (on the same line)?

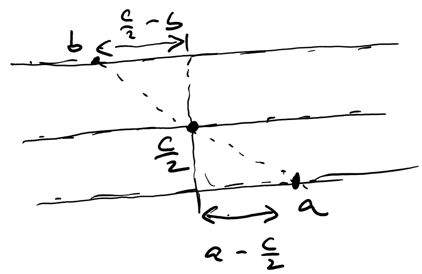


3sum $\leq$ GeomBase

A, B, C
$a+b=c$

$$A \longrightarrow \{(a,0),\ a \in A\}$$
$$B \longrightarrow \{(b,2),\ b \in B\}$$
$$C \longrightarrow \{(\tfrac{c}{2},1),\ c \in C\}$$

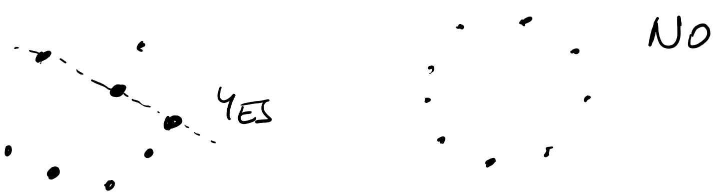
$$a+b=c \iff a - \frac{c}{2} = \frac{c}{2} - b$$



$a+b=c$
iff $(b,2), (\frac{c}{2},1), (a,0)$
are on a line $\blacksquare$

Colinearity:

Given n points in a plane, are any three of them on the same line?



3sum $\leq$ Colinearity:

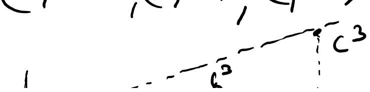
X
$\exists a,b,c \in X$
$a+b+c=0$

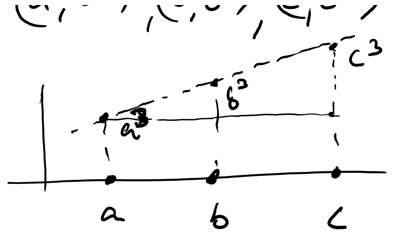
$$X \to (a, a^3) \quad a \in X$$

$a \neq b \neq c$

$(a,a^3), (b,b^3), (c,c^3)$ colinear $\iff a+b+c=0$

$c^3 - a^3 \qquad b^3 - a^3$





$$\frac{c^3-a^3}{c-a} = \frac{b^3-a^3}{b-a} \qquad c \neq a,\ c \neq b$$

$$\Updownarrow$$

$$c^2+ac+a^2 = b^2+ba+a^2$$

$(c^2+ac+a^2)(c-a) = \Rightarrow$

$= c^3 + \cancel{ac^2} + \cancel{a^2c} - \cancel{c^2a} - \cancel{a^2c} - a^3$

$$\Updownarrow$$

$$c^2-b^2+ac-ab=0$$

$$\Updownarrow$$

$$(c+b)(c-b)+a(c-b)=0$$

$$\Updownarrow$$

$$(c+b+a)(c-b)=0$$

$c \neq b$

$$\Updownarrow$$

$$c+b+a=0$$

- For 3Sum, one can first test whether $\exists$ solution of the form $2a+c=0$ (in time $O(n \log n)$) and only then reduce to colinearity.
  Hence, we are interested only in solutions where $a \neq b \neq c$.

$\Rightarrow$ $O(n^{2-\varepsilon})$ alg for GeomBase or Colinearity
$\Rightarrow$ $O(n^{2-\varepsilon})$ for 3Sum.

- GeomBase & Colinearity can be solved in time $\approx O(n^2)$.

$A, B, C,\ x+y=z$

- 3Sum $\leq$ 3Sum with universe size $U = O(n^3)$

$\rightarrow$ hashing. Want: $h : [U] \to [R]$ $\qquad R = O(n^3)$

s.t. $\forall x,y,z \in X \quad h(x)+h(y)=h(z) \iff x+y=z$.

consider the usual hash fcn: $\qquad R \dots$ prime

$$h_{s,t}(x) = sx+t \mod R$$

R … p

$$h_{s,t}(x) = sx + t \mod R$$

$\forall x \neq y \in [U]$ $\quad \Pr_{s,t \in \{1,\dots,R-1\}} [\, h_{s,t}(x) = h_{s,t}(y) \,] = \frac{1}{R}$ $\quad$ (well known fact)

$$\Rightarrow \quad \Pr_{s \in \{1,\dots,R-1\}} [\, h_{s,0}(x) = h_{s,0}(y) \,] = \frac{1}{R}$$

$\rightarrow$ def. $\quad h_s(x) = sx \mod R \quad = h_{s,0}(x)$

$$x + y = z \quad \Rightarrow \quad h_s(x) + h_s(y) = \overbrace{sx \bmod R + sy \bmod R}^{\in [0,\dots,2R-1]}$$

$$h_s(z) = h_s(x+y) = (sx + sy) \bmod R = (sx \bmod R + sy \bmod R) \bmod R$$

$$\Rightarrow h_s(x) + h_s(y) = \begin{cases} h_s(x+y) & \text{if } h_s(x) + h_s(y) < R \\ h_s(x+y) + R & \text{if } h_s(x) + h_s(y) \geq R \end{cases}$$

3sum $\leq$ 3sum $\quad U = O(n^3) \qquad R \approx 6n^3 \quad R \dots$ prime

$$A, B, C \longrightarrow \underset{\substack{\| \\ \{h_s(a),\, a \in A\}}}{h_s(A)},\ h_s(B),\ h_s(C) \cup R + h_s(C)$$

$\exists\, a + b = c$ with $a \in A$, $b \in B$, $c \in C$ $\quad \Longrightarrow \quad \exists\, a' \in h(A)\ \exists\, b' \in h_s(B)\ \exists\, c' \in h_s(C) \cup R + h_s(C)$

$$a' + b' = c'$$

<u>False negative?</u> $\quad a + b \neq c \quad \& \quad h_s(a) + h_s(b) = h_s(c) + \{0, R\}$

$$\Updownarrow$$

$$h_s(a+b) = h_s(c)$$

$$\Pr_s [\, h_s(a+b) = h_s(c) \,] = \frac{1}{R} \quad \text{for } a + b \neq c$$

$$\Rightarrow \Pr_s \left[\exists\, (a,b,c) \in A \times B \times C \ \text{s.t.}\ a + b \neq c \ \&\ h_s(a) + h_s(b) = h_s(c) + \{0, R\} \right]$$

$$\leq \underset{\#\text{triple } a,b,c}{n^3} \cdot \frac{1}{R} \approx \frac{1}{6}$$

R is a prime $\approx 6n^3$.

$\Rightarrow$

1) If there is a solution for A, B, C $\Rightarrow$ $\exists$ is a solution for $h_s(A), h_s(B)$, $h_s(C) \cup R + h_s(C)$

2) If there is no solution for A, B, C $\Rightarrow$ w. prob. $\geq \frac{5}{6}$ over choice of $\underline{s}$, no solution for $h_s(A), h_s(B)$, $h_s(C) \cup R + h_s(C)$

Alg. for 3-Sum on universe $U = O(n^3)$ $O(n^{2-\varepsilon})$ $\Rightarrow$ ~~probabilistic~~ alg. for 3-Sum on arbitrary universe $O(n^{2-\varepsilon})$, prob. of error $\leq \frac{1}{4}$.

Convolution 3SUM

Given an array $A[1, \dots, n]$ are there indices i, j from $\{1, \dots, n\}$ s.t.

$$A[i] + A[j] = A[i+j] \ ?$$

$\rightarrow$ Alg in $O(n^2)$ time - try all possible (i, j).

- $\underline{\text{3Sum} \leq \text{Convolution 3-Sum}}$

X

$\exists\, x + y = z$?

$\underline{\text{reduction}}$: 1) check for solution $2x = z$? $O(n \lg n)$

2) $h : [U] \rightarrow [R]$ $\quad R = n^{1 - \varepsilon/4}$

hash X into buckets $B_1, \dots B_R$

$x \rightarrow B_{h(x)}$

3) $\forall B_i$ where $|B_i| \geq \frac{3n}{R}$, check

$\exists\, x \in B_i,\ y, z \in X$ s.t. $x + y = z$.

(The expected # of items x in all B_i, $|B_i| \geq \frac{3n}{R}$ is $O(R)$ $\Rightarrow$ $O(R \cdot n)$ time in expectation)

see below
$\underline{\text{Lemma 1}}$

Discard those B_i, $|B_i| \geq \frac{3n}{R}$.

4) For each triple $i, j, k \in \left[\frac{3n}{R}\right]$

- set A of size n to $10u$
- put i^{th} elt of B_t to $A[8t+1]$
  put j^{th} elt of B_t to $A[8t+3]$
  put k^{th} elt of B_t to $A[8t+4]$
  for $t \in [1, \dots, R]$.

solve convolution 3sum on A
– if solution found $\Rightarrow$ report.

# of instances of conv. 3sum $\leq \left(\frac{3n}{R}\right)^3 = 27 \cdot n^{\frac{3}{4}\varepsilon}$

If there is an algorithm for Convolution 3sum running in time $O(n^{2-\varepsilon}) \Rightarrow$ alg for 3sum in time $O(n^{2-\frac{\varepsilon}{4}})$.

- If there is a solution to the original 3sum instances, the elts are mapped into some buckets and they will form a solution either in step 3) or 4) of the algorithm.
- Any solution the algorithm reports is clearly a solution to the original problem.
  ($\underset{\substack{\shortparallel \\ x}}{A[i]} + \underset{\substack{\shortparallel \\ y}}{A[j]} = \underset{\substack{\shortparallel \\ z}}{A[i+j]}$ so $x+y=z$)

<u>Lemma 1</u>: If $h: [U] \to [R]$ is picked at random from a hash system so that $\forall x, y \in U$ & $S \subseteq [U]$, $|S| = n$

$\Pr_h[h(x) = h(y)] \leq \frac{1}{R}$ then the expected number of items $\in S$ which have more than $(s-1)$ collisions $\in S$ (are in buckets of size $\geq s$) is $\leq \frac{n}{s - \frac{2n}{R}}$.

$\Rightarrow$ For $s = \frac{3n}{R}$, the expected # of items in buckets of size $\geq \frac{3n}{R}$ is $\leq R$.

Pf: For $h: [\mathcal{U}] \to [R]$ define

$$Q_h = \{ (x,y) \in S \times S ;\ h(x) = h(y) \ \& \ x \neq y \}$$

$$\Rightarrow \quad E_h[|Q_h|] = \sum_{\substack{(x,y) \in \mathcal{U}^2 \\ x \neq y}}' \Pr_h[h(x) = h(y)] \leq \frac{n(n-1)}{R}$$

large ... $\quad L_h = \{ x \in S ;\ |B_h(x)| \geq s \}$ $\qquad B_h(x) = \{ y \in S ;\ h(x) = h(y) \}$, $x \in B_h(x)$

small ... $\quad S_h = \{ x \in S ;\ |B_h(x)| < s \}$

For $i \in R$

$$B_i = \begin{cases} \{ x \in S ;\ h(x) = i \} & |\{ x \in S ;\ h(x) = i \}| < s \\ \emptyset & \text{otherwise} \end{cases}$$

$$S_h = \bigcup_{i \in [R]} B_i$$

$$|Q_h| \geq |L_h| \cdot (s-1) + \underbrace{\sum_{i=1}^{R} |B_i| \cdot (|B_i| - 1)}_{\text{minimal when all } B_i\text{'s are of the same size conditioned on } \sum' |B_i| = |S_h|}$$

$\Downarrow$ $|B_i| = \frac{|S_h|}{R}$

$$\geq |L_h| \cdot (s-1) + |S_h| \cdot \left(\frac{|S_h|}{R} - 1 \right)$$

$$= |L_h| \cdot (s-1) + (n - |L_h|) \cdot \left(\frac{n - |L_h|}{R} - 1 \right)$$

$$\geq |L_h| \cdot (s-1) + |L_h| - n + \frac{n^2 - 2|L_h| n}{R}$$

$$= |L_h| \cdot \left(s - \frac{2n}{R} \right) - n + \frac{n^2}{R}$$

$$\frac{n \cdot (n-1)}{R} \geq \mathbb{E}[|Q_h|] \geq \mathbb{E}[L_h] \cdot \left(s - \frac{2n}{R} \right) - n + \frac{n^2}{R}$$

$$\frac{n \cdot (n-1)}{R} = \frac{n^2}{R} - \frac{n}{R}$$

$$\Downarrow$$

$$n - \frac{n}{R} \geq \mathbb{E}[|L_h|] \left(s - \frac{2n}{R} \right)$$

$$\Rightarrow \quad \frac{n}{\left(s - \frac{2n}{R} \right)} \geq \underbrace{\mathbb{E}[|L_h|]}_{\text{the expected}}$$

$\boxtimes$

$\left(S - \frac{\ }{R}\right)$

the expected
# of elts in buckets
of size $\geq S$.

$\boxed{3}$

Note: Typical bucket has $\frac{n}{R}$ items in expectation.
Total # of items in buckets of size $\geq \frac{3n}{R}$ is $\leq R$
in expectation!
(a bit counterintuitive for small R)

Exact Triangle Problem

- Given a graph G on n vertices with edge weights, is there a triple of vertices a, b, c s.t. $w(a,b) + w(b,c) + w(a,c) = 0$?

$\rightarrow O(n^3)$ algorithm Q: can you do better?

$O(n^{3-\varepsilon})$ alg. for Exact $\triangle$ Problem $\Rightarrow$
$O(n^{2-\frac{\varepsilon}{2}})$ alg for 3Sum

Pf: reduce 3Sum-Convolution for array $\underline{A}$ of size $\underline{n}$ to $\sqrt{n}$ instances of exact triangle on tripartite graphs with $\sqrt{n}$ vertices:

$\forall i \in [0, \sqrt{n}] \longrightarrow G_i = V_i \cup U_i \cup W_i$

$V_i = [0, \sqrt{n}]$, $U_i = [0, 2\sqrt{n}]$, $W_i = [0, \sqrt{n}]$

1) $\forall g \in U_i, \forall t \in V_i$ edge $w(g,t) = A[i\sqrt{n} + (g-t)]$

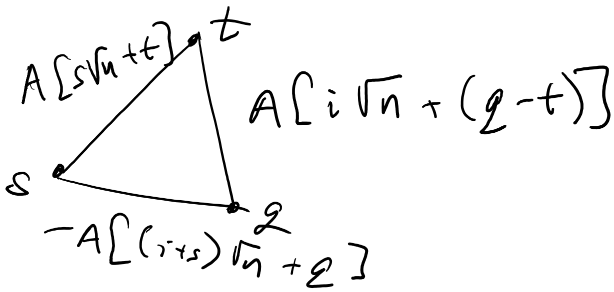
$g - t \in [0, \sqrt{n}-1]$

2) $\forall s \in W_i, t \in V_i$ edge $w(s,t) = A[s\sqrt{n} + t]$

3) $\forall$ ... edge $w(\ ,\ ) = -A[(i+s)\sqrt{n} + g]$

2) $\forall s \in W_i,\ t \in V_i$ edge $w(s,t) = A[s\sqrt{n}+t]$

3) $\forall q \in U_i,\ s \in W_i$; edge $w(q,s) = -A[(i+s)\sqrt{n}+q]$



3sum Solution $A[i\sqrt{n}+(q-t)] + A[s\sqrt{n}+t] = A[(i+s)\sqrt{n}+q]$

$\Longleftrightarrow$

gives a triangle $w(q,t) + w(t,s) + w(q,s) = 0$ in G_i $\blacksquare$

reduction produces $G_1 \dots G_{\sqrt{n}}$ in time $O(n^{3/2})$

solving each instance G_i in time $O((\sqrt{n})^{3-\varepsilon})$

$\Rightarrow$ 3sum-conv in time $O(\sqrt{n} \cdot n^{\frac{3-\varepsilon}{2}})$

$= O(n^{2-\frac{\varepsilon}{2}})$ $\blacksquare$